Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE @ DIRECT®

Theoretical Computer Science 299 (2003) 327–346

Theoretical
Computer Science

www.elsevier.com/locate/tcs

On omega context free languages which are Borel sets of infinite rank

Olivier Finkel

Equipe de Logique Mathématique CNRS et Université Paris 7, U.F.R. de Mathématiques 2 Place Jussieu 75251 Paris cedex 05, France

Received 1 May 2001; received in revised form 15 March 2002; accepted 26 March 2002
Communicated by D. Perrin

Abstract

This paper is a continuation of the study of topological properties of omega context free languages (ω-CFL). We proved in (*Topological properties of omega context free languages*, Theoretical Computer Science, 262 (1–2) (2001) 669–697) that the class of ω-CFL exhausts the finite ranks of the Borel hierarchy, and in (Borel hierarchy and omega context free languages, Theoretical Computer Science, to appear) that there exist some ω-CFL which are analytic but non Borel sets. We prove here that there exist some omega context free languages which are Borel sets of infinite (but not finite) rank, giving additional answer to questions of Lescow and Thomas [Logical specifications of infinite computations in: "A Decade of Concurrency" (J.W. de Bakker et al. (Eds.), Springer LNCS 803 (1994) 583–621).

Keywords: Omega context free languages; Topological properties; Borel hierarchy

1. Introduction

Since J.R. Büchi studied the ω-languages recognized by finite automata to prove the decidability of the monadic second order theory of one successor over the integers [5] the so-called ω-regular languages have been intensively studied. See [43,36] for many results and references.

Pushdown automata are a natural extension of finite automata. Cohen and Gold [9,10] and Linna [29] studied the ω-languages accepted by omega pushdown automata,

E-mail address: finkel@logique.jussieu.fr (O. Finkel).

PII: S0304-3975(02)00327-4

considering various acceptance conditions for omega words. It turned out that the omega languages accepted by omega pushdown automata were also those generated by context free grammars where infinite derivations are considered, also studied by Nivat [33,34] and Boasson and Nivat [4]. These languages were then called the omega context free languages (ω-CFL). See also Staiger's paper [40] for a survey of general theory of ω-languages, including more powerful accepting devices, like Turing machines, and the fundamental study of Engelfriet and Hoogeboom on **X**-automata, i.e. finite automata equipped with a storage type **X**, reading infinite words [18].

Topology is a useful tool for classifying ω-languages by the study of their complexity, particularly with regard to the Borel hierarchy.

McNaughton's Theorem implies that ω-regular languages (ω-languages accepted by deterministic Muller automata) are boolean combination of $\mathbf{\Pi}^0_2$-sets [31]. Topological properties of ω-regular languages were first studied by Landweber in [27] where he characterized ω-regular languages in a given Borel class.

Engelfriet and Hoogeboom proved that all ω-languages accepted by *deterministic* **X**-automata with a Muller acceptance condition are also boolean combinations of $\mathbf{\Pi}^0_2$-sets hence $(\mathbf{\Sigma}^0_3 \cap \mathbf{\Pi}^0_3)$-sets.

When considering *nondeterministic* finite machines, as **X**-automata, a natural question, posed by Lescow and Thomas in [28], now arises: what is the topological complexity of ω-languages accepted by automata equipped with a given storage type **X**? Are they all Borel sets of finite rank, Borel sets, analytic sets?

It is well known that every ω-language accepted by a Turing machine (hence also by a **X**-automaton) with a Muller acceptance condition is an analytic set [40] (i.e. is obtained as a continuous image of a Borel set or as the projection of a Borel set [32]).

We consider in this paper the storage type "*pushdown*". We pursue the investigation of topological properties of omega context free languages. We proved that the class of ω-CFL exhausts the finite ranks of the Borel hierarchy, giving examples of $\mathbf{\Pi}^0_n$-complete (respectively $\mathbf{\Sigma}^0_n$-complete) ω-CFL for each integer $n \geqslant 1$, [20]. We showed in [19] that there exist some omega context free languages which are analytic but non Borel sets. There exist such ω-languages in the form L^ω, with L a context free finitary language; this gave an answer to questions of Niwinski and Simonnet about omega powers of finitary languages [35,38].

But the question was still open whether there exist some omega context free languages which are Borel sets of infinite rank.

We answer to this question in this paper giving examples of ω-CFL which are Borel sets of infinite rank.

The paper is organized as follows. In Sections 2 and 3, we first review some above definitions and results about ω-regular, ω-context free languages, and topology.

In Section 4 we introduce the operation of exponentiation of sets defined by Duparc in his recent study of the Wadge Hierarchy of Borel sets, which is a great refinement of the Borel hierarchy [15], and recall preceding results of [20].

In Section 5, we prove our main result about ω-CFL, using an iteration of Duparc's operation and give additional answer to questions of Thomas and Lescow [28].

2. ω-regular and ω-context free languages

We assume the reader to be familiar with the theory of formal languages and of ω-regular languages, see for example [24,43]. We first recall some of the definitions and results concerning ω-regular and ω-context free languages and omega pushdown automata as presented in [43,9,10].

When Σ is a finite alphabet, a finite string (word) over Σ is any sequence $x = x_1 \ldots x_k$, where $x_i \in \Sigma$ for $i = 1, \ldots, k$ and k is an integer $\geqslant 1$. The length of x is k, denoted by $|x|$.

If $|x| = 0$, x is the empty word denoted by λ.

we write $x(i) = x_i$ and $x[i] = x(1) \ldots x(i)$ for $i \leqslant k$ and $x[0] = \lambda$.

$\Sigma^\star$ is the set of finite words over Σ.

The first infinite ordinal is ω.

An ω-word over Σ is an ω-sequence $a_1 \ldots a_n \ldots$, where $a_i \in \Sigma, \forall i \geqslant 1$.

When σ is an ω-word over Σ, we write $\sigma = \sigma(1)\sigma(2) \ldots \sigma(n) \ldots$

and $\sigma[n] = \sigma(1)\sigma(2) \ldots \sigma(n)$ the finite word of length n, prefix of σ.

The set of ω-words over the alphabet Σ is denoted by Σ^ω.

An ω-language over an alphabet Σ is a subset of Σ^ω.

The usual concatenation product of two finite words u and v is denoted $u.v$ (and sometimes just uv). This product is extended to the product of a finite word u and an ω-word v: the infinite word $u.v$ is then the ω-word such that:

$(u.v)(k) = u(k)$ if $k \leqslant |u|$, and

$(u.v)(k) = v(k - |u|)$ if $k > |u|$.

For $V \subseteq \Sigma^\star$, $V^\omega = \{\sigma = u_1 \ldots u_n \ldots \in \Sigma^\omega / u_i \in V, \forall i \geqslant 1\}$ is the ω-power of V.

For $V \subseteq \Sigma^\star$, the complement of V (in $\Sigma^\star$) is $\Sigma^\star - V$ denoted V^-.

For a subset $A \subseteq \Sigma^\omega$, the complement of A is $\Sigma^\omega - A$ denoted A^-.

The prefix relation is denoted $\sqsubseteq$: the finite word u is a prefix of the finite word v (denoted $u \sqsubseteq v$) if and only if there exists a (finite) word w such that $v = u.w$.

This definition is extended to finite words which are prefixes of ω-words:

the finite word u is a prefix of the ω-word v (denoted $u \sqsubseteq v$) iff there exists an ω-word w such that $v = u.w$.

Definition 2.1. A finite state machine (FSM) is a quadruple $M = (K, \Sigma, \delta, q_0)$, where K is a finite set of states, Σ is a finite input alphabet, $q_0 \in K$ is the initial state and δ is a mapping from $K \times \Sigma$ into 2^K. A FSM is called deterministic (DFSM) iff: $\delta : K \times \Sigma \to K$.
A Büchi automaton (BA) is a 5-tuple $M = (K, \Sigma, \delta, q_0, F)$ where $M' = (K, \Sigma, \delta, q_0)$ is a finite state machine and $F \subseteq K$ is the set of final states.
A Muller automaton (MA) is a 5-tuple $M = (K, \Sigma, \delta, q_0, \mathscr{F})$ where $M' = (K, \Sigma, \delta, q_0)$ is a FSM and $\mathscr{F} \subseteq 2^K$ is the collection of designated state sets.
A Büchi or Muller automaton is said deterministic if the associated FSM is deterministic.
Let $\sigma = a_1 a_2 \ldots a_n \ldots$ be an ω-word over Σ.
A sequence of states $r = q_1 q_2 \ldots q_n \ldots$ is called an (infinite) run of $M = (K, \Sigma, \delta, q_0)$ on σ, starting in state p, iff: (1) $q_1 = p$ and (2) for each $i \geqslant 1$, $q_{i+1} \in \delta(q_i, a_i)$.

In case a run r of M on σ starts in state q_0, we call it simply "a run of M on σ".
For every (infinite) run $r = q_1 q_2 \ldots q_n \ldots$ of M, $In(r)$ is the set of states in K entered by M infinitely many times during run r:
$In(r) = \{q \in K / \{i \geqslant 1 / q_i = q\}$ is infinite$\}$.
For $M = (K, \Sigma, \delta, q_0, F)$ a BA, the ω-language accepted by M is $L(M) = \{\sigma \in \Sigma^\omega /$there exists a run r of M on σ such that $In(r) \cap F \neq \emptyset\}$.
For $M = (K, \Sigma, \delta, q_0, \mathscr{F})$ a MA, the ω-language accepted by M is $L(M) = \{\sigma \in \Sigma^\omega /$there exists a run r of M on σ such that $In(r) \in \mathscr{F}\}$.

The classical result of Mc Naughton [31] established that the expressive power of deterministic MA (DMA) is equal to the expressive power of non deterministic MA (NDMA) which is also equal to the expressive power of non deterministic BA (NDBA).

There is also a characterization of the languages accepted by MA by means of the "ω-Kleene closure" which we give now the definition:

Definition 2.2. For any family L of finitary languages over the alphabet Σ, the ω-Kleene closure of L, is

$$\omega\text{-}KC(L) = \left\{ \bigcup_{i=1}^{n} U_i.V_i^\omega / U_i, V_i \in L, \forall i \in [1,n] \right\}.$$

Theorem 2.3. *For any ω-language L, the following conditions are equivalent*:
(1) *L belongs to ω-$KC(REG)$, where REG is the class of (finitary) regular languages.*
(2) *There exists a DMA that accepts L.*
(3) *There exists a MA that accepts L.*
(4) *There exists a BA that accepts L.*
An ω-language L satisfying one of the conditions of the above theorem is called an ω-regular language.

The class of ω-regular languages will be denoted by REG_ω.

We now define the pushdown machines and the classes of ω-context free languages.

Definition 2.4. A pushdown machine (PDM) is a 6-tuple $M = (K, \Sigma, \Gamma, \delta, q_0, Z_0)$, where K is a finite set of states, Σ is a finite input alphabet, Γ is a finite pushdown alphabet, $q_0 \in K$ is the initial state, $Z_0 \in \Gamma$ is the start symbol, and δ is a mapping from $K \times (\Sigma \cup \{\lambda\}) \times \Gamma$ to finite subsets of $K \times \Gamma^\star$.
If $\gamma \in \Gamma^+$ describes the pushdown store content, the leftmost symbol will be assumed to be on "top" of the store. A configuration of a PDM is a pair (q, γ) where $q \in K$ and $\gamma \in \Gamma^\star$.
For $a \in \Sigma \cup \{\lambda\}$, $\gamma, \beta \in \Gamma^\star$ and $Z \in \Gamma$, if (p, β) is in $\delta(q, a, Z)$, then we write $a : (q, Z\gamma) \mapsto_M (p, \beta\gamma)$.
$\mapsto_M^\star$ is the transitive and reflexive closure of $\mapsto_M$. (The subscript M will be omitted whenever the meaning remains clear).

Let $\sigma = a_1a_2\ldots a_n\ldots$ be an ω-word over Σ. An infinite sequence of configurations $r=(q_i,\gamma_i)_{i\geqslant 1}$ is called a run of M on σ, starting in configuration (p,γ), iff:

(1) $(q_1,\gamma_1)=(p,\gamma)$

(2) for each $i\geqslant 1$, there exists $b_i\in\Sigma\cup\{\lambda\}$ satisfying $b_i : (q_i,\gamma_i)\mapsto_M (q_{i+1},\gamma_{i+1})$ such that either $a_1a_2\ldots a_n\ldots = b_1b_2\ldots b_n\ldots$
or $b_1b_2\ldots b_n\ldots$ is a finite prefix of $a_1a_2\ldots a_n\ldots$

The run r is said to be complete when $a_1a_2\ldots a_n\ldots = b_1b_2\ldots b_n\ldots$
As for FSM, for every such run, $In(r)$ is the set of all states entered infinitely often during run r.
A complete run r of M on σ, starting in configuration (q_0,Z_0), will be simply called "a run of M on σ".

Definition 2.5. A Büchi pushdown automaton (BPDA) is a 7-tuple $M=(K,\Sigma,\Gamma,\delta,q_0,Z_0,F)$ where $M'=(K,\Sigma,\Gamma,\delta,q_0,Z_0)$ is a PDM and $F\subseteq K$ is the set of final states.
The ω-language accepted by M is $L(M)=\{\sigma\in\Sigma^\omega$/there exists a complete run r of M on σ such that $In(r)\cap F\neq\emptyset\}$.

Definition 2.6. A Muller pushdown automaton (MPDA) is a 7-tuple $M=(K,\Sigma,\Gamma,\delta,q_0,Z_0,\mathscr{F})$ where $M'=(K,\Sigma,\Gamma,\delta,q_0,Z_0)$ is a PDM and $\mathscr{F}\subseteq 2^K$ is the collection of designated state sets.
The ω-language accepted by M is $L(M)=\{\sigma\in\Sigma^\omega$/there exists a complete run r of M on σ such that $In(r)\in\mathscr{F}\}$.

Remark 2.7. We consider here two acceptance conditions for ω-words, the Büchi and the Muller acceptance conditions, respectively denoted 2-acceptance and 3-acceptance in [27] and in [10] and $(inf,\sqcap)$ and $(inf,=)$ in [40].

Cohen and Gold, and independently Linna, established a characterization Theorem for ω-*CFL*:

Theorem 2.8. *Let CFL be the class of context free (finitary) languages. Then for any ω-language L the following three conditions are equivalent*:

(1) $L\in\omega$-*KC*(*CFL*).

(2) *There exists a BPDA that accepts L.*

(3) *There exists a MPDA that accepts L.*

In [9] are also studied the ω-languages generated by ω-context free grammars and it is shown that each of conditions (1), (2), and (3) of the above theorem is also equivalent to: (4) L is generated by a context free grammar G by leftmost derivations. These grammars are also studied in [33,34]

Then we can set the following definition:

Definition 2.9. An ω-language is an ω-context free language (ω-*CFL*) iff it satisfies one of the conditions of the above theorem.

3. Topology

We assume the reader to be familiar with basic notions of topology which may be found in [32,28,40,36] and with the elementary theory of (countable) ordinals.

Topology is an important tool for the study of ω-languages, and leads to characterization of several classes of ω-languages.

For a finite alphabet X, we consider X^ω as a topological space with the Cantor topology. The open sets of X^ω are the sets in the form $W.X^\omega$, where $W \subseteq X^\star$. A set $L \subseteq X^\omega$ is a closed set iff its complement $X^\omega - L$ is an open set. The class of open sets of X^ω will be denoted by $\mathbf{G}$ or by $\mathbf{\Sigma_1^0}$. The class of closed sets will be denoted by $\mathbf{F}$ or by $\mathbf{\Pi_1^0}$. Closed sets are characterized by the following:

Proposition 3.1. *A set $L \subseteq X^\omega$ is a closed set of X^ω iff for every $\sigma \in X^\omega$,*
$[\forall n \geqslant 1, \exists u \in X^\omega$ such that $\sigma(1)\ldots\sigma(n).u \in L]$ implies that $\sigma \in L$.

Define now the next classes of the Borel Hierarchy:

Definition 3.2. The classes $\mathbf{\Sigma_n^0}$ and $\mathbf{\Pi_n^0}$ of the Borel Hierarchy on the topological space X^ω are defined as follows:
$\mathbf{\Sigma_1^0}$ is the class of open sets of X^ω.
$\mathbf{\Pi_1^0}$ is the class of closed sets of X^ω.
$\mathbf{\Pi_2^0}$ or $\mathbf{G_\delta}$ is the class of countable intersections of open sets of X^ω.
$\mathbf{\Sigma_2^0}$ or $\mathbf{F_\sigma}$ is the class of countable unions of closed sets of X^ω.
And for any integer $n \geqslant 1$:
$\mathbf{\Sigma_{n+1}^0}$ is the class of countable unions of $\mathbf{\Pi_n^0}$-subsets of X^ω.
$\mathbf{\Pi_{n+1}^0}$ is the class of countable intersections of $\mathbf{\Sigma_n^0}$-subsets of X^ω.
The Borel Hierarchy is also defined for transfinite levels. The classes $\mathbf{\Sigma_\alpha^0}$ and $\mathbf{\Pi_\alpha^0}$, for a countable ordinal α, are defined in the following way:
$\mathbf{\Sigma_\alpha^0}$ is the class of countable unions of subsets of X^ω in $\bigcup_{\gamma<\alpha} \mathbf{\Pi_\gamma^0}$.
$\mathbf{\Pi_\alpha^0}$ is the class of countable intersections of subsets of X^ω in $\bigcup_{\gamma<\alpha} \mathbf{\Sigma_\gamma^0}$.

Recall some basic results about these classes [32]:

Proposition 3.3. (a) $\mathbf{\Sigma_\alpha^0} \cup \mathbf{\Pi_\alpha^0} \subsetneq \mathbf{\Sigma_{\alpha+1}^0} \cap \mathbf{\Pi_{\alpha+1}^0}$, *for each countable ordinal* $\alpha \geqslant 1$.
(b) $\bigcup_{\gamma<\alpha} \mathbf{\Sigma_\gamma^0} = \bigcup_{\gamma<\alpha} \mathbf{\Pi_\gamma^0} \subsetneq \mathbf{\Sigma_\alpha^0} \cap \mathbf{\Pi_\alpha^0}$, *for each countable limit ordinal* α.
(c) *A set $W \subseteq X^\omega$ is in the class $\mathbf{\Sigma_\alpha^0}$ iff its complement is in the class $\mathbf{\Pi_\alpha^0}$.*
(d) $\mathbf{\Sigma_\alpha^0} - \mathbf{\Pi_\alpha^0} \neq \emptyset$ *and* $\mathbf{\Pi_\alpha^0} - \mathbf{\Sigma_\alpha^0} \neq \emptyset$ *hold for every countable ordinal* $\alpha \geqslant 1$.
(e) *For every ordinal $\alpha \geqslant 1$, the class $\mathbf{\Sigma_\alpha^0}$ is closed under countable unions and the class $\mathbf{\Pi_\alpha^0}$ is closed under countable intersections.*

We shall say that a subset of X^ω is a Borel set of rank α, for a countable ordinal α, iff it is in $\mathbf{\Sigma_\alpha^0} \cup \mathbf{\Pi_\alpha^0}$ but not in $\bigcup_{\gamma<\alpha}(\mathbf{\Sigma_\gamma^0} \cup \mathbf{\Pi_\gamma^0})$.

There is a nice characterization of $\mathbf{\Pi_2^0}$-subsets of X^ω. First define the notion of W^δ:

Definition 3.4. For $W \subseteq X^\star$, let:
$W^\delta = \{\sigma \in X^\omega / \exists^\omega i$ such that $\sigma[i] \in W\}$.
($\sigma \in W^\delta$ iff σ has infinitely many prefixes in W).

Then we can state the following Proposition:

Proposition 3.5 (see Staiger [40]). *A subset L of X^ω is a $\mathbf{\Pi^0_2}$-subset of X^ω iff there exists a set $W \subseteq X^\star$ such that $L = W^\delta$.*

For X a finite set, (and this is also true if X is an infinite alphabet) there are some subsets of X^ω which are not Borel sets. Indeed there exists another hierarchy beyond the Borel hierarchy, which is called the projective hierarchy and which is obtained from the Borel hierarchy by successive applications of operations of projection and complementation. More precisely, a subset A of X^ω is in the class $\mathbf{\Sigma^1_1}$ of **analytic** sets iff there exists another finite set Y and a Borel subset B of $(X \times Y)^\omega$ such that $x \in A \leftrightarrow \exists y \in Y^\omega$ such that $(x, y) \in B$.
We denote (x, y) the infinite word over the alphabet $X \times Y$ such that $(x, y)(i) = (x(i), y(i))$ for each integer $i \geqslant 0$.
Now a subset of X^ω is in the class $\mathbf{\Pi^1_1}$ of **coanalytic** sets iff its complement in X^ω is an analytic set.
The next classes are defined in the same manner, $\mathbf{\Sigma^1_{n+1}}$-sets of X^ω are projections of $\mathbf{\Pi^1_n}$-sets and $\mathbf{\Pi^1_{n+1}}$-sets are the complements of $\mathbf{\Sigma^1_{n+1}}$-sets.
Recall also the notion of completeness with regard to reduction by continuous functions. Let α be a countable ordinal. A set $F \subseteq X^\omega$ is a $\mathbf{\Sigma^0_\alpha}$ (respectively, $\mathbf{\Pi^0_\alpha}$)-complete set iff for any set $E \subseteq Y^\omega$ (with Y a finite alphabet):
$E \in \mathbf{\Sigma^0_\alpha}$ (respectively, $E \in \mathbf{\Pi^0_\alpha}$) iff there exists a continuous function $f : Y^\omega \to X^\omega$ such that $E = f^{-1}(F)$.
A similar notion exists for the classes of the projective hierarchy: in particular A set $F \subseteq X^\omega$ is a $\mathbf{\Sigma^1_1}$ (respectively, $\mathbf{\Pi^1_1}$)-complete set iff for any set $E \subseteq Y^\omega$ (Y a finite alphabet):
$E \in \mathbf{\Sigma^1_1}$ (respectively, $E \in \mathbf{\Pi^1_1}$) iff there exists a continuous function f such that $E = f^{-1}(F)$.
A $\mathbf{\Sigma^0_\alpha}$ (respectively, $\mathbf{\Pi^0_\alpha}$)-complete set is a $\mathbf{\Sigma^0_\alpha}$ (respectively, $\mathbf{\Pi^0_\alpha}$)-set which is in some sense a set of the highest topological complexity among the $\mathbf{\Sigma^0_\alpha}$ (respectively, $\mathbf{\Pi^0_\alpha}$)-sets. $\mathbf{\Sigma^0_n}$ (respectively $\mathbf{\Pi^0_n}$)-complete sets, with n an integer $\geqslant 1$, are thoroughly characterized in [39].
Landweber studied first the topological properties of ω-regular languages. He proved that every ω-regular language is a boolean combination of $\mathbf{G_\delta}$-sets. and he also characterized the ω-regular languages in each of the Borel classes $\mathbf{F}, \mathbf{G}, \mathbf{F_\sigma}, \mathbf{G_\delta}$, and showed that one can decide, for an effectively given ω-regular language L, whether L is in $\mathbf{F}, \mathbf{G}, \mathbf{F_\sigma}$, or $\mathbf{G_\delta}$.
It turned out that an ω-regular language is in the class $\mathbf{G}_\delta$ iff it is accepted by a deterministic Büchi automaton.

When considering ω-*CFL*, natural questions now arise: are all ω-*CFL* Borel sets of finite rank, Borel sets, analytic sets....?

First recall the following previous result [40]:

Theorem 3.6. *Every ω-CFL over a finite alphabet X is an analytic subset of X^ω.*

We showed the following

Theorem 3.7 (Finkel [19]). (a) *There exist some ω-CFL which are $\mathbf{\Sigma^1_1}$-complete sets hence non Borel sets.*

(b) *It is undecidable whether an effectively given ω-CFL is a Borel set.*

Next the ω-*CFL* exhaust the finite ranks of the Borel hierarchy.

Theorem 3.8 (Finkel [20]). *For each nonnegative integer $n \geqslant 1$, there exist $\mathbf{\Sigma^0_n}$-complete ω-CFL A_n* and $\mathbf{\Pi^0_n}$-complete *ω-CFL B_n.*

Cohen and Gold proved that one cannot decide whether an ω-*CFL* is in the class $\mathbf{F}, \mathbf{G}$ or $\mathbf{G_\delta}$. We have extended in [20] this result to all classes $\mathbf{\Sigma^0_n}$ and $\mathbf{\Pi^0_n}$, for n an integer $\geqslant 1$, and next to all Borel classes in [19]. (We say that an ω-*CFL* A is effectively given when a MPDA accepting A is given.)

But the question was still open whether there exist some omega context free languages which are Borel sets of infinite (but not finite) rank. We shall show below that there exist such omega context free languages.

4. Operation "exponentiation of sets"

In order to construct omega context free languages of every finite rank, we used recent results of Duparc about the Wadge hierarchy. The Wadge hierarchy of Borel sets is a huge refinement of the Borel hierarchy. Wadge gave first a description of this hierarchy [44] and Duparc recently got a new proof of Wadge's results and he gave a normal form of Borel sets, i.e. an inductive construction of a Borel set of every given degree [13,15]. In fact we shall need in this paper only some of his results. So we shall recall only these results and refer to [13,15] for more details.

Duparc's proof relies on set theoretic operations which are the counterpart of arithmetical operations over ordinals needed to compute the Wadge degrees. In fact we shall only use in this paper the operation of exponentiation over sets of *infinite* words. Moreover we shall consider a slight modification of Duparc's operation $A \to A^\sim$ we introduced in [20] and which we recall now:

Definition 4.1. Let X_A be a finite alphabet and $\twoheadleftarrow \notin X_A$.
Let $X = X_A \cup \{\twoheadleftarrow\}$ and x be a finite or infinite word over the alphabet X.
Then $x^{\twoheadleftarrow}$ is inductively defined by
$\lambda^{\twoheadleftarrow} = \lambda$,
for a finite word $u \in (X_A \cup \{\twoheadleftarrow\})^\star$:

$(u.a)^{\twoheadleftarrow} = u^{\twoheadleftarrow}.a$, if $a \in X_A$,
$(u. \twoheadleftarrow)^{\twoheadleftarrow} = u^{\twoheadleftarrow}$ with its last letter removed if $|u^{\twoheadleftarrow}| > 0$,
$(u. \twoheadleftarrow)^{\twoheadleftarrow}$ is undefined if $|u^{\twoheadleftarrow}| = 0$,
and for u infinite:
$(u)^{\twoheadleftarrow} = \lim_{n\in\omega}(u[n])^{\twoheadleftarrow}$, where, given β_n and v in $X_A^{\star}$,
$v \sqsubseteq \lim_{n\in\omega} \beta_n \leftrightarrow \exists n \forall p \geqslant n \quad \beta_p[|v|] = v$.

Remark 4.2. For $x \in X^{\leqslant\omega}$, $x^{\twoheadleftarrow}$ denotes the string x, once every $\twoheadleftarrow$ occuring in x has been "evaluated" to the back space operation (the one familiar to your computer!), proceeding from left to right inside x. In other words $x^{\twoheadleftarrow} = x$ from which every interval of the form "$a \twoheadleftarrow$" ($a \in X_A$) is removed. We add the convention that $(u. \twoheadleftarrow)^{\twoheadleftarrow}$ is undefined if $|u^{\twoheadleftarrow}| = 0$, i.e. when the last letter $\twoheadleftarrow$ cannot be used as an eraser (because every letter of X_A in u has already been erased by some erasers $\twoheadleftarrow$ placed in u). Remark that the resulting word $x^{\twoheadleftarrow}$ may be finite or infinite.

For example if $u = (a \twoheadleftarrow)^n$, for $n \geqslant 1$, $u = (a \twoheadleftarrow)^\omega$ or $u = (a \twoheadleftarrow\twoheadleftarrow)^\omega$ then $(u)^{\twoheadleftarrow} = \lambda$,
if $u = (ab \twoheadleftarrow)^\omega$ then $(u)^{\twoheadleftarrow} = a^\omega$,
if $u = bb(\twoheadleftarrow a)^\omega$ then $(u)^{\twoheadleftarrow} = b$,
if $u = \twoheadleftarrow (a \twoheadleftarrow)^\omega$ or $u = a \twoheadleftarrow\twoheadleftarrow a^\omega$ then $(u)^{\twoheadleftarrow}$ is undefined.
We can now define the variant $A \to A^{\approx}$ of the operation of exponentiation of sets:

Definition 4.3. For $A \subseteq X_A^\omega$ and $\twoheadleftarrow \notin X_A$, let $X = X_A \cup \{\twoheadleftarrow\}$ and
$A^{\approx} = \{x \in (X_A \cup \{\twoheadleftarrow\})^\omega / x^{\twoheadleftarrow} \in A\}$.

The following result is then another formulation of a property of the operation $A \to A^{\sim}$ proved in [15] and which was applied in [20] to study the ω-powers of finitary context free languages.

Theorem 4.4. *Let n be an integer $\geqslant 2$ and $A \subseteq X_A^\omega$ be a $\mathbf{\Pi^0_n}$-complete set. Then $A^{\approx}$ is a $\mathbf{\Pi^0_{n+1}}$-complete subset of $(X_A \cup \{\twoheadleftarrow\})^\omega$.*

We proved that the class CFL_ω is closed under this operation $A \to A^{\approx}$.

Theorem 4.5 (Finkel [20]). *Whenever $A \subseteq X_A^\omega$ is an ω-CFL, then $A^{\approx} \subseteq (X_A \cup \{\twoheadleftarrow\})^\omega$ is an ω-CFL.*

Proof. An ω-word $\sigma \in A^{\approx}$ may be considered as an ω-word $\sigma^{\twoheadleftarrow} \in A$ to which we possibly add, before the first letter $\sigma^{\twoheadleftarrow}(1)$ of $\sigma^{\twoheadleftarrow}$ (respectively between two consecutive letters $\sigma^{\twoheadleftarrow}(n)$ and $\sigma^{\twoheadleftarrow}(n+1)$ of $\sigma^{\twoheadleftarrow}$), a finite word v_1 (respectively v_{n+1}) where:
for all integers $i \geqslant 1$, v_i belongs to the context free (finitary) language L_3 generated by the context free grammar with the following production rules:
$S \to aS \twoheadleftarrow S$ with $a \in X_A$,
$S \to \lambda$ (λ being the empty word).

This language L_3 corresponds to words where every letter of X_A has been removed after using the back space operation.

Remark 4.6. Recall that a one counter automaton is a pushdown automaton with a pushdown alphabet in the form $\Gamma=\{Z_0, z\}$ where Z_0 is the bottom symbol and always remains at the bottom of the pushdown store. And a one counter language is a (finitary) language which is accepted by a one counter automaton by final states. It is easy to see that in fact L_3 is a deterministic one-counter language, i.e. L_3 is accepted by a deterministic one-counter automaton. And for $a\in X_A$, the language $L_3.a$ is also accepted by a deterministic one-counter automaton.

Then we can see that whenever $A\subseteq X_A^\omega$, the ω-language $A^{\approx}\subseteq(X_A\cup\{\twoheadleftarrow\})^\omega$ is obtained by substituting in A the language $L_3.a$ for each letter $a\in X_A$, where L_3 is the *CFL* defined above.

Let now A be an ω-*CFL* given by $A=\bigcup_{i=1}^{n} U_i.V_i^\omega$ where U_i and V_i are context free languages. Then $A^{\approx}=\bigcup_{i=1}^{n}(U_i').V_i'^\omega$, where U_i' (respectively V_i') is obtained by substituting the language $L_3.a$ to each letter $a\in X_A$ in U_i (respectively V_i).

The class *CFL* is closed under substitution, so U_i' and V_i' are *CFL*. Hence the ω-language $A^{\approx}$ is an ω-*CFL* because ω-$KC(CFL)\subseteq CFL_\omega$.

We have also given in [20] an effective construction of a MPDA accepting the ω-language $A^{\approx}\subseteq(X_A\cup\{\twoheadleftarrow\})^\omega$ from a MPDA accepting an ω-language $A\subseteq X_A^\omega$. Recall now the idea of this construction.

Let A be an ω-CFL which is accepted by a Muller pushdown automaton $\mathcal{A}=(K, X_A, \Gamma, \delta, q_0, Z_0, \mathcal{F})$. The ω-language accepted by $\mathcal{A}$ is $L(\mathcal{A})=A=\{\sigma\in X_A^\omega/\text{there exists a run } r \text{ of } \mathcal{A} \text{ on } \sigma \text{ such that } In(r)\in\mathcal{F}\}$.

We can construct another MPDA $\mathcal{A}^{\approx}$ which accepts the ω-language $A^{\approx}$ over the alphabet $X=X_A\cup\{\twoheadleftarrow\}$.

Let us describe informally the behaviour of the machine $\mathcal{A}^{\approx}$ when it reads an ω-word $\sigma\in A^{\approx}$. Recall that this word may be considered as an ω-word $\sigma^{\twoheadleftarrow}\in A$ to which we possibly add, before the first letter $\sigma^{\twoheadleftarrow}(1)$ of $\sigma^{\twoheadleftarrow}$ (respectively between two consecutive letters $\sigma^{\twoheadleftarrow}(n)$ and $\sigma^{\twoheadleftarrow}(n+1)$ of $\sigma^{\twoheadleftarrow}$), a finite word v_1 (respectively v_{n+1}) where v_i belongs to the context free language L_3.

$\mathcal{A}^{\approx}$ starts the reading as a pushdown automaton accepting the language L_3. Then $\mathcal{A}^{\approx}$ begins to read as $\mathcal{A}$, but at any moment of the computation it may guess (using the nondeterminism) that it reads a finite segment v of L_3 which will be erased (using the eraser $\twoheadleftarrow$). It reads v using an additional stack letter E which permits to simulate a one counter automaton at the top of the stack while keeping the memory of the stack of $\mathcal{A}$. Then, after the reading of v, $\mathcal{A}^{\approx}$ simulates again the machine $\mathcal{A}$ and so on.

5. ω-*CFL* which are Borel of infinite rank

A well known example of $\mathbf{\Pi}_2^0$-complete ω-regular language is

$B_2=\{\alpha\in\{0,1\}^\omega/\exists^\omega i\ \alpha(i)=1\}=(0^\star.1)^\omega$,

where $\exists^\omega i$ means: "there exist infinitely many i such that ...".

B_2 is an omega context free language because it is an ω-regular language.

We can now get some $\mathbf{\Pi^0_{n+2}}$-complete set, for an integer $n \geqslant 1$, from the $\mathbf{\Pi^0_2}$-complete set B_2 by applying $n \geqslant 1$ times the operation of exponentiation of sets.

More precisely, we define, for a set $A \subseteq X_A^\omega$:
$A^{\approx.0} = A$
$A^{\approx.1} = A^{\approx}$ and
$A^{\approx.(k+1)} = (A^{\approx.k})^{\approx}$,
where we apply $k+1$ times the operation $A \to A^{\approx}$ with different new letters $\twoheadleftarrow_1$, $\twoheadleftarrow_2$, $\twoheadleftarrow_3, \dots, \twoheadleftarrow_{k+1}$.

We can now infer from Theorems 4.4 and 4.5 that, for an integer $n \geqslant 1$, $(B_2)^{\approx.n}$ is an omega context free language which is a $\mathbf{\Pi^0_{n+2}}$-complete subset of $\{0, 1, \twoheadleftarrow_1, \dots, \twoheadleftarrow_n\}^\omega$. Similarly, if $A \subseteq X_A^\omega$ is a $\mathbf{\Pi^0_2}$-complete regular or context free ω-language over the alphabet X_A, the ω-language $(A)^{\approx.n}$ is a $\mathbf{\Pi^0_{n+2}}$-complete subset of $(X_A \cup \{\twoheadleftarrow_1, \dots, \twoheadleftarrow_n\})^\omega$.

A way to obtain a Borel set of infinite rank, as we shall show below, is to define, for two letters a, b in X_A, the supremum of the sets $A^{\approx.i}$:

$$\sup_{i \in \mathbb{N}} A^{\approx.i} = \bigcup_{i \in \mathbb{N}} a^i.b.A^{\approx.i}.$$

But this set is defined over an infinite alphabet, and any omega context free language is defined over a finite alphabet. So we have first to code this set over a finite alphabet. We shall first code every set $A^{\approx.n}$. The ω-language $A^{\approx.n}$ is defined over the alphabet $X_A \cup \{\twoheadleftarrow_1, \dots, \twoheadleftarrow_n\}$ hence we have to code every eraser $\twoheadleftarrow_j$ by a finite word over a fixed finite alphabet. We shall code the eraser $\twoheadleftarrow_j$ by the finite word $\alpha.B^j.C^j.D^j.E^j.\beta$ over the alphabet $\{\alpha, B, C, D, E, \beta\}$. The reason of the coding we choose will be clear later, when we construct a Muller pushdown automaton accepting an ω-language close to the coding of $\sup_{i \in \mathbb{N}} A^{\approx.i}$. In fact this MPDA needs to read four times the integer j characterizing the eraser $\twoheadleftarrow_j$.

Remark first that one can define the morphism

$$F_n : (X_A \cup \{\twoheadleftarrow_1, \dots, \twoheadleftarrow_n\})^\star \to (X_A \cup \{\alpha, \beta, B, C, D, E\})^\star$$

by $F(c) = c$ for each $c \in X_A$ and $F(\twoheadleftarrow_j) = \alpha.B^j.C^j.D^j.E^j.\beta$ for each integer $j \in [1, n]$, where $B, C, D, E, \alpha, \beta$ are new letters not in X_A. This morphism is naturally extended to a continuous function

$$\bar{F}_n : (X_A \cup \{\twoheadleftarrow_1, \dots, \twoheadleftarrow_n\})^\omega \to (X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega.$$

Then $\bar{F}_n((X_A \cup \{\twoheadleftarrow_1, \dots, \twoheadleftarrow_n\})^\omega)$ is the continuous image by $\bar{F}_n$ of the compact set $(X_A \cup \{\twoheadleftarrow_1, \dots, \twoheadleftarrow_n\})^\omega$, hence it is also a compact set, and a closed subset of $(X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega$. We can now state the following lemma. Its proof is easy and left to the reader.

Lemma 5.1. *Let $A \subseteq X_A^\omega$ be a $\mathbf{\Pi^0_2}$-complete subset of X_A^ω. Then for each integer $n \geqslant 1$, the ω-language $\bar{F}_n(A^{\approx.n})$ is a $\mathbf{\Pi^0_{n+2}}$-complete subset of $(X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega$.*

We shall prove now that the supremum of the sets $\bar{F}_n(A^{\approx.n})$ is a Borel set of infinite rank.

Lemma 5.2. *Let $A \subseteq X_A^\omega$ be a $\mathbf{\Pi_2^0}$-complete subset of X_A^ω. Then the set*

$$\sup_{n\geqslant 1} \bar{F}_n(A^{\approx.n}) = \bigcup_{n\geqslant 1} a^n.b.\bar{F}_n(A^{\approx.n})$$

is a $\mathbf{\Sigma_\omega^0}$-subset of $(X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega$ which is not a Borel set of finite rank.

Proof. Assume $A \subseteq X_A^\omega$ is $\mathbf{\Pi_2^0}$-complete. Then the preceding lemma implies that, for each $n \geqslant 1$, the ω-language $\bar{F}_n(A^{\approx.n})$ is a $\mathbf{\Pi_{n+2}^0}$-complete subset of $(X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega$. Let a, b be two letters in X_A then it is easy to show that, for each $n \geqslant 1$, the set $a^n.b.\bar{F}_n(A^{\approx.n})$ is also a $\mathbf{\Pi_{n+2}^0}$-complete subset of $(X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega$ thus

$$\sup_{n\geqslant 1} \bar{F}_n(A^{\approx.n}) = \bigcup_{n\geqslant 1} a^n.b.\bar{F}_n(A^{\approx.n})$$

is in the class $\mathbf{\Sigma_\omega^0}$ by definition of this class.

On the other side this set cannot be a Borel set of finite rank. Because if $\sup_{n\geqslant 1} \bar{F}_n(A^{\approx.n})$ was in the class $\mathbf{\Pi_j^0}$, for an integer $j \geqslant 1$, then the set

$$a^n.b.\bar{F}_n(A^{\approx.n}) = \sup_{n\geqslant 1} \bar{F}_n(A^{\approx.n}) \cap (a^n.b.(X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega)$$

would be also in the class $\mathbf{\Pi_j^0}$, because $(a^n.b.(X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega)$ is a closed hence $\mathbf{\Pi_j^0}$-set and the class of $\mathbf{\Pi_j^0}$-subsets of $(X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega$ is closed under finite intersection. But this would lead to a contradiction because we have seen that, for $n \geqslant j$, the set $a^n.b.\bar{F}_n(A^{\approx.n})$ is $\mathbf{\Pi_{n+2}^0}$-complete, where $n+2 \geqslant j+2 > j$ hence it is not a $\mathbf{\Pi_j^0}$-subset of $(X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega$. $\square$

We cannot show that the ω-language $\sup_{n\geqslant 1} \bar{F}_n(A^{\approx.n})$ is an omega context free language. This is connected to the fact that the finitary language

$$\{B^j C^j D^j E^j / j \geqslant 1\}$$

is not a context free language. But its complement is easily seen to be context free. Then, instead of considering $\sup_{n\geqslant 1} \bar{F}_n(A^{\approx.n})$, we can add to this ω-language all ω-words in the form $a^n.b.u$ where there is in u a segment $\alpha.B^j.C^k.D^l.E^m.\beta$, with j, k, l, m integers $\geqslant 1$, which does not code any eraser, or codes an eraser $\twoheadleftarrow_j$ for $j > n$. Then we add to $\sup_{n\geqslant 1} \bar{F}_n(A^{\approx.n})$ another ω-language which is of Borel rank 2 and the resulting ω-language will be still of infinite rank, but we shall show that it is an omega context free language.

So we define now formally this construction in the following way.

Define first the following context free finitary languages over the alphabet $X^{\square}=(X_A\cup\{\alpha,\beta,B,C,D,E\})$:

$$L^B = \{a^n.b.u.B^j | n \geqslant 1 \text{ and } j > n \text{ and } u \in (X^{\square})^{\star}\},$$

$$L^C = \{a^n.b.u.C^j | n \geqslant 1 \text{ and } j > n \text{ and } u \in (X^{\square})^{\star}\},$$

$$L^D = \{a^n.b.u.D^j | n \geqslant 1 \text{ and } j > n \text{ and } u \in (X^{\square})^{\star}\},$$

$$L^E = \{a^n.b.u.E^j | n \geqslant 1 \text{ and } j > n \text{ and } u \in (X^{\square})^{\star}\},$$

$$L^{(B,C)} = \{u.\alpha.B^j.C^k.D^l.E^m.\beta | j,k,l,m \geqslant 1 \text{ and } j \neq k \text{ and } u \in (X^{\square})^{\star}\},$$

$$L^{(C,D)} = \{u.\alpha.B^j.C^k.D^l.E^m.\beta | j,k,l,m \geqslant 1 \text{ and } k \neq l \text{ and } u \in (X^{\square})^{\star}\},$$

$$L^{(D,E)} = \{u.\alpha.B^j.C^k.D^l.E^m.\beta | j,k,l,m \geqslant 1 \text{ and } l \neq m \text{ and } u \in (X^{\square})^{\star}\}.$$

Let now

$$L = L^B \cup L^C \cup L^D \cup L^E \cup L^{(B,C)} \cup L^{(C,D)} \cup L^{(D,E)}.$$

It is easy to show that each of the languages $L^B, L^C, L^D, L^E, L^{(B,C)}, L^{(C,D)}, L^{(D,E)}$ is a context free finitary language thus L is also context free because the class CFL is closed under finite union. Then the ω-language $L.(X^{\square})^{\omega}$ is an ω-CFL which is an open subset of $(X^{\square})^{\omega}$.

Remark now that any word in $\sup_{n\geqslant 1} \bar{F}_n(A^{\approx.n})$ belongs to the regular ω-language

$$R = a^+.b.(X_A \cup (\alpha.B^+.C^+.D^+.E^+.\beta))^{\omega}$$

because every word has an initial segment in the form $a^n.b$ with $n\geqslant 1$ and the letters $\alpha, B, C, D, E, \beta$ are only used to code the erasers $\twoheadleftarrow_j$ for $j\geqslant 1$.

Consider now the ω-language

$$L.(X^{\square})^{\omega} \cap R.$$

An ω-word σ in this language is a word in R such that σ has an initial word in the form $a^n.b$, with $n\geqslant 1$, and σ contains a segment $\alpha.B^j.C^k.D^l.E^m.\beta$ with $j,k,l,m\geqslant 1$ which does not code any eraser $\twoheadleftarrow_j$ or codes such an eraser but with $j>n$. Thus this ω-language is disjoint from the set $\sup_{n\geqslant 1} \bar{F}_n(A^{\approx.n})$. Consider now the ω-language:

$$A^{\bullet} = \sup_{n\geqslant 1} \bar{F}_n(A^{\approx.n}) \cup [L.(X^{\square})^{\omega} \cap R].$$

We can now state the next lemma.

Lemma 5.3. *Let $A\subseteq X_A^{\omega}$ be a $\mathbf{\Pi}^0_2$-complete subset of X_A^{ω}. Then $A^{\bullet}$ is a $\mathbf{\Sigma}^0_{\boldsymbol{\omega}}$-subset of $(X_A\cup\{\alpha,\beta,B,C,D,E\})^{\omega}$ which is not a Borel set of finite rank.*

Proof. Let $A\subseteq X_A^{\omega}$ be a $\mathbf{\Pi}^0_2$-complete subset of X_A^{ω}. Then we have already seen that $\sup_{n\geqslant 1} \bar{F}_n(A^{\approx.n})$ is a $\mathbf{\Sigma}^0_{\boldsymbol{\omega}}$-subset of $(X_A\cup\{\alpha,\beta,B,C,D,E\})^{\omega}$. On the other side it is easy to see, from Proposition 3.5, that the ω-regular language R is a $\mathbf{\Pi}^0_2$-set because

$$R = (R')^{\delta},$$

where R' is the finitary (regular) language defined by

$$R' = a^+.b.(X_A \cup (\alpha.B^+.C^+.D^+.E^+.\beta))^+.$$

Then the ω-language

$$L.(X^{\square})^\omega \cap R$$

is the intersection of an open set and of a $\mathbf{\Pi^0_2}$-set. Thus it is also a $\mathbf{\Pi^0_2}$-set because the class $\mathbf{\Pi^0_2}$ is closed under finite intersection. Then the ω-language

$$A^\bullet = \sup_{n\geqslant 1} \bar{F}_n(A^{\approx .n}) \cup [L.(X^{\square})^\omega \cap R]$$

is a $\mathbf{\Sigma^0_\omega}$-subset of $(X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega$ because the class $\mathbf{\Sigma^0_\omega}$ is closed under finite union.

We want now to prove that $A^\bullet$ is not a Borel set of finite rank. Assume, on the contrary, that $A^\bullet$ is of finite rank J, where J is an integer $\geqslant 1$. Then the intersection of $A^\bullet$ and of the complement of $L.(X^{\square})^\omega \cap R$ would be the intersection of a $\mathbf{\Pi^0_{J+1}}$-set and of a $\mathbf{\Sigma^0_2}$ hence $\mathbf{\Pi^0_3}$-set. Hence

$$\sup_{n\geqslant 1} \bar{F}_n(A^{\approx .n}) = A^\bullet \cap (L.(X^{\square})^\omega \cap R)^-$$

would be a $\mathbf{\Pi^0_k}$-set, with $k = \max(3, J+1)$. But this is not possible because we know from the preceding lemma that $\sup_{n\geqslant 1} \bar{F}_n(A^{\approx .n})$ is a Borel set of infinite rank. $\square$

We can now state the following

Theorem 5.4. *Let $A \subseteq X_A^\omega$ be an ω-regular language over the alphabet X_A. Then the ω-language*

$$A^\bullet = \sup_{n\geqslant 1} \bar{F}_n(A^{\approx .n}) \cup [L.(X^{\square})^\omega \cap R]$$

is an ω-CFL over the alphabet $(X_A \cup \{\alpha, \beta, B, C, D, E\})$.

Proof. We have already seen that $L.(X^{\square})^\omega$ is an ω-*CFL*, thus

$$L.(X^{\square})^\omega \cap R$$

is also an ω-*CFL* because the class of omega context free languages is closed under intersection with ω-regular languages [9].

Suppose the ω-regular language $A \subseteq X_A^\omega$ is accepted by the deterministic Muller automaton $\mathscr{A} = (K, X_A, \delta, q_0, \mathscr{F}_{\mathscr{A}})$ where $\mathscr{A}' = (K, X_A, \delta, q_0)$ is a FSM and $\mathscr{F}_{\mathscr{A}} \subseteq 2^K$ is the collection of designated state sets.

We shall find a MPDA $\mathscr{B}$ accepting an ω-CFL $L(\mathscr{B})$ such that

$$\sup_{n\geqslant 1} \bar{F}_n(A^{\approx \cdot n}) \subseteq L(\mathscr{B}) \subseteq A^\bullet = \sup_{n\geqslant 1} \bar{F}_n(A^{\approx \cdot n}) \cup [L.(X^{\square})^\omega \cap R].$$

Thus we shall have

$$A^\bullet = L(\mathscr{B}) \cup [L.(X^{\square})^\omega \cap R],$$

and this will imply that $A^\bullet$ is an ω-CFL because the class CFL_ω is closed under finite union [9].

It is easy to have $L(\mathscr{B}) \subseteq R$ because if $L(\mathscr{B}')$ is an ω-CFL which is not included into R one can replace it by $L(\mathscr{B}) = L(\mathscr{B}') \cap R$ which is then an ω-CFL verifying $L(\mathscr{B}) = L(\mathscr{B}') \cap R \subseteq R$.

Recall now that

$$L.(X^{\square})^\omega \cap R$$

is the set of *all* ω-words in R having an initial segment in the form $a^n.b$, with $n \geqslant 1$, and containing a segment $\alpha.B^j.C^k.D^l.E^m.\beta$ with $j,k,l,m \geqslant 1$ which does not code any eraser $\twoheadleftarrow_j$ or codes such an eraser but with $j > n$.

Thus, in order to define the MPDA $\mathscr{B}$, we have only to consider the behaviour of $\mathscr{B}$ when reading ω-words in the form

$$a^n.b.u,$$

where $n \geqslant 1$ and $u \in (X_A \cup (\alpha.B^+.C^+.D^+.E^+.\beta))^\omega$ is such that the letters $\alpha, B, C, D, E, \beta$ in u are only used to code the erasers $\twoheadleftarrow_j$ for $1 \leqslant j \leqslant n$. (In order to simplify our notations, we shall sometimes write in the sequel $\twoheadleftarrow_j = \alpha.B^j.C^j.D^j.E^j.\beta$ and call eraser either $\twoheadleftarrow_j$ or its code $\alpha.B^j.C^j.D^j.E^j.\beta$, with $j \geqslant 1$).

And we have to find a MPDA $\mathscr{B}$ such that $L(\mathscr{B})$ contains such a word $a^n.b.u$ if and only if $u \in \bar{F}_n(A^{\approx.n})$.

So we have to look first at ω-words in $\bar{F}_n(A^{\approx.n})$. In such a word $\sigma \in \bar{F}_n(A^{\approx.n})$, there are (codes of) erasers $\twoheadleftarrow_1, \ldots, \twoheadleftarrow_n$. The ω-word σ is in $\bar{F}_n(A^{\approx.n})$ if and only if after the operations of erasing (with the erasers $\twoheadleftarrow_1, \ldots, \twoheadleftarrow_n$) have been achieved in σ, then the resulting word is in A.

Because of the inductive definition of the sets $A^{\approx.n}$, the operations of erasing have to be done in a good order: in an ω-word which contains only the erasers $\twoheadleftarrow_1, \ldots, \twoheadleftarrow_n$, the first operation of erasing uses the last eraser $\twoheadleftarrow_n$, then the second one uses the eraser $\twoheadleftarrow_{n-1}$, and so on …

Therefore these operations satisfy the following properties:

(a) An eraser $\twoheadleftarrow_j$ may only erase letters $c \in X_A$ or other erasers $\twoheadleftarrow_k$ with $k < j$.

(b) Assume that in a word $u \in \bar{F}_n(A^{\approx.n})$, there is a segment $c.v.x$ where c is either in X_A or in the set $\{\twoheadleftarrow_1, \ldots, \twoheadleftarrow_{n-1}\}$, and x is (the code of) an eraser $\twoheadleftarrow_k$ which erases c when the operations of erasing are successively achieved. Now if there is in the segment v (the code of) an eraser $\twoheadleftarrow_j$ which erases e, where $e \in X_A$ or e is (the code of) another eraser, then e must belong to v (it is between c and x in the word u); moreover the operation of erasing using the eraser $\twoheadleftarrow_j$ has been achieved before that one using the eraser $\twoheadleftarrow_k$ and this implies that $k \leqslant j$. Thus the integer k must verify

$$k \leqslant \min[j/\text{an eraser } \twoheadleftarrow_j \text{ has been used into } v].$$

We can now informally describe the behaviour of the MPDA $\mathscr{B}$ when reading a word $a^n.b.u$ such that the letters $\alpha, B, C, D, E, \beta$ are only used in u to code the erasers $\twoheadleftarrow_j$ for $1 \leqslant j \leqslant n$.

After the reading of the initial segment in the form $a^n.b$, the MPDA $\mathscr{B}$ simulates the Muller automaton $\mathscr{A}$ until it guesses, using the non determinism, that it begins to read a segment w which contains erasers which really erase and some letters of X_A or some other erasers which are erased when the operations of erasing are achieved in u.

Then, using the nondeterminism, when $\mathscr{B}$ reads a letter $c \in X_A$ it may guess that this letter will be erased and push it in the pushdown store, keeping in memory the current state of the Muller automaton $\mathscr{A}$.

In a similar manner when $\mathscr{B}$ reads the code $\twoheadleftarrow_j = \alpha.B^j.C^j.D^j.E^j.\beta$ of an eraser, it may guess that this eraser will be erased (by another eraser $\twoheadleftarrow_k$ with $k > j$) and then it pushes in the store the finite word $\gamma.E^j.\varepsilon$, where γ, E, ε are in the pushdown alphabet.

But $\mathscr{B}$ may also guess that the eraser $\twoheadleftarrow_j = \alpha.B^j.C^j.D^j.E^j.\beta$ will really be used as an eraser. If it guesses that the code of $\twoheadleftarrow_j$ will be used as an eraser, $\mathscr{B}$ has to pop from the top of the pushdown store either a letter $c \in X_A$ or the code $\gamma.E^i.\varepsilon$ of another eraser $\twoheadleftarrow_i$, with $i < j$, which is erased by $\twoheadleftarrow_j$.

It would be easy for $\mathscr{B}$ to check whether $i < j$ when reading the initial segment $\alpha.B^j$ of $\twoheadleftarrow_j$.

But as we remarked in item (b) above, the MPDA $\mathscr{B}$ has also to check that the integer j is smaller or equal than every integer p such that an eraser $\twoheadleftarrow_p$ has been used since the letter $c \in X_A$ or the code $\gamma.E^i.\varepsilon$ was pushed in the store. Then, after having pushed in the pushdown store some letter $x \in X_A$ or the code $x = \gamma.E^i.\varepsilon$ of an eraser, and before it pops it from the top of the store, $\mathscr{B}$ has to keep in the memory of the stack the integer

$$k = \min[p/\text{some eraser } \twoheadleftarrow_p \text{ has been used since } x \text{ was pushed in the stack}].$$

For that purpose $\mathscr{B}$ pushes the finite word $L_2.S^k.L_1$ in the pushdown store (L_1 is pushed first, then S^k and the letter L_2 are pushed in the stack), where L_1, L_2 and S are new letters added to the pushdown alphabet.

So, when $\mathscr{B}$ guesses that $\twoheadleftarrow_j = \alpha.B^j.C^j.D^j.E^j.\beta$ will be really used as an eraser, there is at the top of the stack either a letter $c \in X_A$ or a code $\gamma.E^i.\varepsilon$ of an eraser which will be erased or a code $L_2.S^k.L_1$. The behaviour of $\mathscr{B}$ is then as follows.

Assume first there is at the top of the stack a code $L_2.S^k.L_1$. Then $\mathscr{B}$ firstly checks that $j \leqslant k$ when reading the segment $\alpha.B^j.C$ of the eraser $\alpha.B^j.C^j.D^j.E^j.\beta$.

If $j \leqslant k$ holds, then $\mathscr{B}$ pops completely, using λ-transitions, the word $L_2.S^k.L_1$ from the top of the stack. ($\mathscr{B}$ has checked it is allowed to use the eraser $\twoheadleftarrow_j$).

Then there is now in every case at the top of the stack either a letter $c \in X_A$ or a code $\gamma.E^i.\varepsilon$ of an eraser which will be erased. The MPDA $\mathscr{B}$ pops this letter c or the code $\gamma.E^i.\varepsilon$ (having checked that $j > i$ after reading the segment $\alpha.B^j.C^j$ of the eraser $\alpha.B^j.C^j.D^j.E^j.\beta$).

We have to consider what is now at the top of the stack and distinguish three cases:

(a) If there is now at the top of the stack the bottom symbol Z_0, then the MPDA $\mathscr{B}$, after having completely read the eraser $\twoheadleftarrow_j$, may pursue the simulation of the Muller automaton $\mathscr{A}$ or guesses that it begins to read another segment v which will be erased, hence the next letter $c \in X_A$ or the next code $\alpha.B^m.C^m.D^m.E^m.\beta$ of the word will be erased and then $\mathscr{B}$ pushes the letter $c \in X_A$ or the code $\gamma.E^m.\varepsilon$ of $\twoheadleftarrow_m$ in the pushdown store.

(2) If there is now at the top of the stack either a letter $c' \in X_A$ or a code $\gamma.E^m.\varepsilon$, then $\mathscr{B}$ pushes the code $L_2.S^j.L_1$ in the pushdown store (j is then the minimum of the set of integers p such that an eraser $\twoheadleftarrow_p$ has been used since the letter c' or the code $\gamma.E^m.\varepsilon$ has been pushed into the stack).

(3) If there is now at the top of the stack a code $L_2.S^l.L_1$, then the MPDA $\mathscr{B}$ has to compare the integers j and l and to replace $L_2.S^l.L_1$ by $L_2.S^j.L_1$ if $j<l$. $\mathscr{B}$ achieves this task while reading the segment $D^j.E^j.\beta$ of the eraser $\alpha.B^j.C^j.D^j.E^j.\beta$. The MPDA $\mathscr{B}$ pops a letter S for each letter D read. It then determines whether $j \leqslant l$.

If $j \leqslant l$ then $\mathscr{B}$ pushes $L_2.S^j.L_1$ when reading the segment $E^j.\beta$ of the eraser $\twoheadleftarrow_j$.

If $j>l$, then when every letter S of the code $L_2.S^l.L_1$ has been popped, there are $(j-l)$ letters D of the eraser $\alpha.B^j.C^j.D^j.E^j.\beta$ which have not yet been read by $\mathscr{B}$. When reading these letters the MPDA $\mathscr{B}$ pushes $(j-l)$ letters U in the stack, where U is a new letter in the pushdown alphabet. Then, when reading the segment E^j of the eraser $\alpha.B^j.C^j.D^j.E^j.\beta$, the MPDA $\mathscr{B}$ firstly pops U^{j-l} (when reading the first $(j-l)$ letters E); afterwards $\mathscr{B}$ pushes again S^l in the stack when reading the rest of the eraser $\alpha.B^j.C^j.D^j.E^j.\beta$.

When the content of the stack is again just Z_0, the initial stack symbol of the MPDA $\mathscr{B}$, then $\mathscr{B}$ may pursue the simulation of the Muller automaton $\mathscr{A}$ or guesses it begin to read a new segment which will be erased when the operations of erasing will be successively achieved. □

We can now state our main result:

Theorem 5.5. *Let $A \subseteq X_A^\omega$ be a $\mathbf{\Pi}_2^0$-complete ω-regular language over the alphabet X_A. Then $A^\bullet$ is an omega context free language which is a $\mathbf{\Sigma}_\omega^0$-subset of $(X_A \cup \{\alpha, \beta, B, C, D, E\})^\omega$* but is not a Borel set of finite rank.

Proof. It follows directly from Lemma 5.3 and Theorem 5.4. □

In particular if $B_2 = \{\alpha \in \{0,1\}^\omega / \exists^\omega i \ \ \alpha(i)=1\} = (0^\star.1)^\omega$, then $(B_2)^\bullet$ is an omega context free language which is a Borel set of infinite rank.

Theorem 5.5 provides infinitely many such *ω-CFL* over any finite alphabet X of cardinal $\geqslant 2$, because there exist infinitely many $\mathbf{\Pi}_2^0$-complete ω-regular languages over the alphabet X, and for such an ω-regular language A it holds that

$$A^\bullet \cap X^\omega = a^+.b.A.$$

6. Concluding remarks and further work

We knew that the class of omega context free languages exhausts the finite ranks of the Borel hierarchy and that there exist some *ω-CFL* which are analytic but non Borel sets. We have proven above that there exist some omega context free languages which are Borel sets of infinite rank.

It is well known that Turing machines, with a Büchi or Muller acceptance condition, accept ω-languages of every Borel rank $<\omega_1^{CK}$, where ω_1^{CK} is the first non recursive ordinal [40,32,38]. Then the following problem naturally arises: describe the set of infinite Borel ranks of omega context free languages, and in particular find the ordinal

$$\sup\{\alpha/\text{there exist some Borel } \omega\text{-}CFL \text{ of rank } \alpha\}$$

which is of course $\leqslant \omega_1^{CK}$. Unfortunately we cannot reach some Borel ranks $>\omega$ by iterating our operation $A \to A^{\bullet}$. In fact one cannot even reach some $\mathbf{\Sigma}^0_{\boldsymbol{\omega}}$-complete set, as it will be explained in [22] by considering the Wadge degrees of Borel sets.

Recall that the Wadge hierarchy of Borel sets is a great refinement of the Borel hierarchy. We proved in [21] that the length of the Wadge hierarchy of Borel ω-CFL is an ordinal greater than or equal to the Cantor ordinal ε_0, which is the first fixed point of the ordinal exponentiation of base ω. Using the above construction of $A^{\bullet}$, we have improved this result, showing that this length is an ordinal greater than or equal to ε_ω, which is the ω^{th} fixed point of the ordinal exponentiation of base ω [22].

Acknowledgements

Thanks to the anonymous referees for useful comments on a previous version of this paper.

References

[1] J-M. Autebert, J. Berstel, L. Boasson, Context free languages and pushdown automata, Handbook of Formal Languages, Vol. 1, Springer, Berlin, 1996.

[2] R. Barua, The Hausdorff–Kuratowski hierarchy of ω-regular languages and a hierarchy of muller automata, Theoret. Comput. Sci. 96 (1992) 345–360.

[3] J. Berstel, Transductions and context free languages, Teubner Studienbücher Informatik, 1979.

[4] L. Boasson, M. Nivat, Adherences of Languages, J. Comput. System Sci. 20 (3) (1980) 285–309.

[5] J.R. Büchi, Weak second order arithmetic and finite automata, Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik 6 (1960) 66–92.

[6] J.R. Büchi, On a decision method in restricted second order arithmetic, logic methodology and philosophy of science, Proc. 1960 Int. Congr., Stanford University Press, 1962, pp. 1–11.

[7] J.R. Büchi, L.H. Landweber, Solving sequential conditions by finite state strategies. Trans. Amer. Math. Soc. 138 (1969) 295–311.

[8] J.R. Büchi, D. Siefkes, The monadic second order theory of all countable ordinals, Decidable Theories 2, Springer Lecture Notes in Mathematics, Vol. 328, Berlin, 1973, pp. 1–217.

[9] R.S. Cohen, A.Y. Gold, Theory of ω-languages, Parts one and two, J. Comput. System Sci., 15 (2) (1977) 169–184; 185–208.

[10] R.S. Cohen, A.Y. Gold, ω-computations on deterministic pushdown machines, J. Comput. System Sci. 16 (3) (1978) 275–300.

[11] O. Carton, D. Perrin, Chains and superchains for ω-rational sets, Automata and semigroups, Internat. J. Algebra Comput. 7 (7) (1997) 673–695.
[12] O. Carton, D. Perrin, The Wagner hierarchy of ω-rational sets, Internat. J. Algebra Comput. 9 (5) (1999) 597–620.
[13] J. Duparc, La Forme Normale des Boréliens de Rang Fini, Ph.D. Thesis, Université Paris 7, 1995.
[14] J. Duparc, The normal form of Borel sets, Part 1: Borel Sets of Finite Rank, C.R.A.S. Paris, t.320, Série 1, 1995, pp. 651–656.
[15] J. Duparc, Wadge Hierarchy and Veblen hierarchy: Part 1: Borel sets of finite rank J. Symbolic Logic 66 (1) (2001) 56–86.
[16] J. Duparc, A hierarchy of context free omega languages, Theoret. Comput. Sci., to appear.
[17] J. Duparc, O. Finkel, J-P. Ressayre, Computer science and the fine structure of Borel sets, Theoret. Comput. Sci. 257 (1–2) (2001) 85–105.
[18] J. Engelfriet, H.J. Hoogeboom, X-automata on ω-words, Theoret. Comput. Sci. 110 (1) (1993) 1–51.
[19] O. Finkel, Borel hierarchy and omega context free languages, Theoret. Comput. Sci., to appear.
[20] O. Finkel, Topological properties of omega context free languages, Theoret. Comput. Sci. 262 (1–2) (2001) 669–697.
[21] O. Finkel, Wadge hierarchy of omega context free languages, Theoret. Comput. Sci. 269 (1–2) (2001) 283–315.
[22] O. Finkel, On the wadge hierarchy of omega context free languages, in: the Proceedings of the International Workshop Logic and Complexity in Computer Science LCCS 2001 held in honour of A. Slissenko for his 60th birthday, Creteil, September 2001, pp. 69–79.
[23] S. Ginsburg, The mathematical theory of context free languages, Mc Graw-Hill Book Company, New York, 1966.
[24] J.E. Hopcroft, J.D. Ullman, Formal languages and their Relation to Automata, Addison-Wesley Publishing Company, Reading, MA, 1969.
[25] M. Kaminsky, A classification of ω-regular languages, Theoret. Comput. Sci. 36 (1985) 217–229.
[26] K. Kuratowski, Topology, Academic Press, New York, 1966.
[27] L.H. Landweber, Decision problems for ω-automata, Math. System Theory 3 (1969) 376–384.
[28] H. Lescow, W. Thomas, Logical specifications of infinite computations, in: J.W. de Bakker, et al., (Eds.), A Decade of Concurrency, Lecture Notes in Computer Science, Vol. 803, Springer, Berlin, 1994, pp. 583–621.
[29] M. Linna, On ω-sets associated with context-free languages, Inform. Control 31 (3) (1976) 272–293.
[30] M. Linna, A decidability result for deterministic ω-context-free languages, Theoret. Comput. Sci. 4 (1977) 83–98.
[31] R. Mac Naughton, Testing and generating infinite sequences by a finite automaton, Inform. Control 9 (1966) 521–530.
[32] Y.N. Moschovakis, Descriptive Set Theory, North-Holland, Amsterdam, 1980.
[33] M. Nivat, Mots Infinis Engendrés par une Grammaire Algébrique, RAIRO Inform. Théory 11 (1977) 311–327.
[34] M. Nivat, Sur les Ensembles de Mots Infinis Engendrés par une Grammaire Algébrique, RAIRO Inform. Théory 12 (1978) 259–278.
[35] D. Niwinski, Problem on ω-powers posed in: the Proceedings of the 1990 Workshop, Logics and Recognizable Sets, Univ. Kiel.
[36] D. Perrin, J.-E. Pin, Infinite words, available from http://www.liafa.jussieu.fr/jep/InfiniteWords.html Book in preparation.
[37] V. Selivanov, Fine hierarchy of regular ω-languages, Theoret. Comput. Sci. 191 (1998) 37–59.
[38] P. Simonnet, Automates et Théorie Descriptive, Ph.D. Thesis, Université Paris 7, March 1992.
[39] L. Staiger, Hierarchies of recursive ω-languages, J. Inform. Process. Cybernetics EIK 22 (5/6) (1986) 219–241.
[40] L. Staiger, ω-languages, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Vol. 3, Springer, Berlin, 1997.
[41] L. Staiger, On ω-power languages, in: New Trends in Formal Languages, Control, Cooperation, and Combinatorics, Lecture Notes in Computer Science, Vol. 1218, Springer, Berlin, 1997, pp. 377–393.

[42] L. Staiger, K. Wagner, Automatentheoretische und Automatenfreie Charakterisierungen Topologischer Klassen Regulärer Folgenmengen, Elektron. Informationsverarb Kybernetik EIK 10 (7) (1974) 379–392.
[43] W. Thomas, Automata on infinite objects, in: J. Van Leeuwen (Ed.), Handbook of Theoretical Computer Science, vol. B, Elsevier, Amsterdam, 1990, pp. 133–191.
[44] W.W. Wadge, Ph.D. Thesis, Berkeley, 1984.
[45] K. Wagner, On omega regular sets, Inform. Control 43 (1979) 123–177.
[46] Th. Wilke, H. Yoo, Computing the wadge degree, the Lifschitz degree and the Rabin index of a regular language of infinite words in polynomial time, in: P.D. Mosses, M. Nielsen, M.I. Schwartzbach (Eds.), TAPSOFT' 95: Theory and Practice of Software Development, Lecture Notes in Computer Science, Vol. 915, Springer, Berlin, 1995, pp. 288–302.